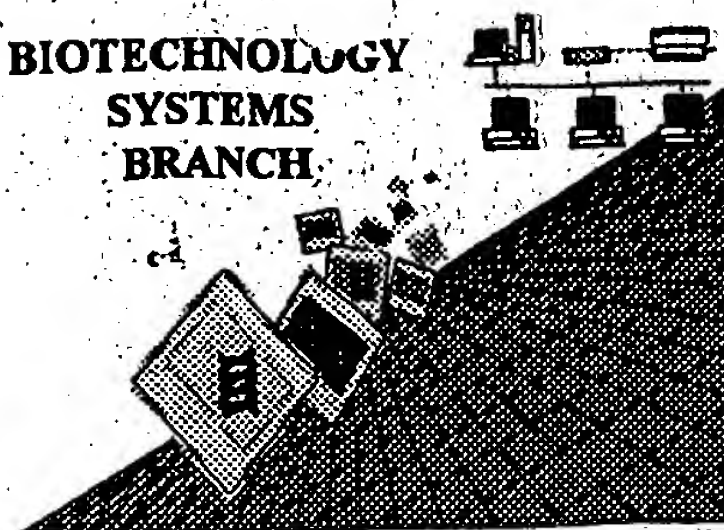


# **RAW SEQUENCE LISTING** **ERROR REPORT**

BIOTECHNOLOGY  
SYSTEMS  
BRANCH



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/866,570

Source: OIPR

Date Processed by STIC: 6/15/2001

**THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.**

**PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:**

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,**
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY**

**FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.**

**FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.**

**PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)**

**TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:**

## **Checker Version 3.0**

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

**Checker Version 3.0 can be down loaded from the USPTO website at the following address:**

**<http://www.uspto.gov/web/offices/pac/checker>**

# Raw Sequence Listing Error Summary

## ERROR DETECTED

## SUGGESTED CORRECTION

SERIAL NUMBER: 09/866,570

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1        Wrapped Nucleics  
    Wrapped Aminos   The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2        Invalid Line Length   The rules require that a line not exceed 72 characters in length. This includes white spaces.
- 3        Misaligned Amino  
    Numbering       The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
- 4        Non-ASCII       The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
- 5        Variable Length   Sequence(s)        contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6        PatentIn 2.0  
    "bug"           A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s)       . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 7        Skipped Sequences  
    (OLD RULES)   Sequence(s)        missing. If intentional, please insert the following lines for each skipped sequence:  
                  (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                  (i)       SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
                  (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
                  This sequence is intentionally skipped  
  
                  Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
- 8        Skipped Sequences  
    (NEW RULES)   Sequence(s)        missing. If intentional, please insert the following lines for each skipped sequence.  
                  <210> sequence id number  
                  <400> sequence id number  
                  000
- 9        Use of n's or Xaa's  
    (NEW RULES)   Use of n's and/or Xaa's have been detected in the Sequence Listing.  
                  Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present.  
                  In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 10        Invalid <213>  
    Response       Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
- 11        Use of <220>       Sequence(s) 28 missing the <220> "Feature" and associated numeric identifiers and responses.  
                  Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
                  (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12        PatentIn 2.0  
    "bug"           Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,570

DATE: 06/15/2001

TIME: 16:35:41

Input Set : A:\53679.APP.txt

Output Set: N:\CRF3\06152001\I866570.raw

pp1, 3-4  
Does Not Comply  
Corrected Diskette Needed

3 <110> APPLICANT: Croteau, Rodney et al.  
5 <120> TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway  
7 <130> FILE REFERENCE: 53679  
C--> 9 <140> CURRENT APPLICATION NUMBER: US/09/866,570  
C--> 10 <141> CURRENT FILING DATE: 2001-05-25  
12 <150> PRIOR APPLICATION NUMBER: 09/411,145  
13 <151> PRIOR FILING DATE: 1999-09-30  
15 <160> NUMBER OF SEQ ID NOS: 58  
17 <170> SOFTWARE: PatentIn Ver. 2.0

## ERRORED SEQUENCES

751 <210> SEQ ID NO: 18  
752 <211> LENGTH: 302  
753 <212> TYPE: PRT  
754 <213> ORGANISM: Taxus cuspidata  
756 <400> SEQUENCE: 18  
757 Phe Tyr Pro Phe Ala Gly Arg Met Arg Asn Lys Gly Asp Gly Glu Leu  
758 1 5 10 15  
760 Glu Val Asp Cys Thr Gly Glu Gly Ala Leu Phe Val Glu Ala Met Ala  
761 20 25 30  
763 Asp Asp Asn Leu Ser Val Leu Gly Gly Phe Asp Tyr His Asn Pro Ala  
764 35 40 45  
766 Phe Gly Lys Leu Leu Tyr Ser Leu Pro Leu Asp Thr Pro Ile His Asp  
767 50 55 60  
769 Leu His Pro Leu Val Val Gln Val Thr Arg Phe Thr Cys Gly Gly Phe  
770 65 70 75 80  
772 Val Val Gly Leu Ser Leu Asp His Thr Ile Cys Asp Gly Arg Gly Ala  
773 85 90 95  
775 Gly Gln Phe Leu Lys Ala Leu Ala Glu Met Ala Arg Gly Glu Ala Lys  
776 100 105 110  
778 Pro Ser Leu Glu Pro Ile Met Asn Arg Glu Leu Leu Lys Pro Glu Asp  
779 115 120 125  
781 Leu Ile Arg Leu Gln Phe Tyr His Phe Glu Ser Met Arg Pro Pro Pro  
782 130 135 140  
784 Ile Val Glu Glu Met Val Gln Ser Ser Ile Ile Ile Asn Ala Glu Thr  
785 145 150 155 160  
E--> 787 Ile Ser Asn Xaa Lys Gln Tyr Ile Met Glu Cys Lys Glu Ser Cys  
788 165 170 175  
790 Ser Ala Phe Asp Val Val Gly Gly Leu Ala Met Leu Ala Arg Thr Lys  
791 180 185 190  
793 Ala Phe Gln Ile Pro His Thr Glu Asn Val Met Val Ile Phe Ala Val  
794 195 200 205  
796 Asp Ala Arg Arg Ser Phe Asp Pro Pro Leu Pro Lys Gly Tyr Tyr Gly  
797 210 215 220  
799 Asn Val Val Gly Asn Ala Cys Ala Leu Asp Asn Val Gln Asp Leu Leu

→ see  
item 9 on  
Error  
summary sheet

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,570

DATE: 06/15/2001

TIME: 16:35:41

Input Set : A:\53679.APP.txt

Output Set: N:\CRF3\06152001\I866570.raw

```

800 225          230          235          240
802 Asn Gly Ser Leu Leu Arg Ala Thr Met Ile Ile Lys Lys Ser Lys Val
803          245          250          255
805 Ser Leu Lys Glu Asn Ile Arg Ala Lys Thr Leu Thr Ile Pro Ser Ile
806          260          265          270
808 Val Asp Val Asn Val Lys His Glu Asn Ile Val Gly Leu Gly Asp Leu
809          275          280          285
811 Arg Arg Leu Gly Phe Asn Glu Val Asp Phe Gly Trp Gly Lys
812          290          295          300
2185 <210> SEQ ID NO: 58
2186 <211> LENGTH: 438
2187 <212> TYPE: PRT
2188 <213> ORGANISM: Taxus cuspidata
2190 <400> SEQUENCE: 58
2191 Met Glu Lys Leu His Val Asp Ile Ile Glu Arg Val Lys Val Ala Pro
2192 1          5          10          15
2194 Cys Leu Pro Ser Ser Lys Glu Ile Leu Gln Leu Ser Ser Leu Asp Asn
2195          20          25          30
2197 Ile Leu Arg Cys Tyr Val Ser Val Leu Phe Val Tyr Asp Arg Val Ser
2198          35          40          45
2200 Thr Val Ser Ala Asn Pro Ala Lys Thr Ile Arg Glu Ala Leu Ser Lys
2201          50          55          60
2203 Val Leu Val Tyr Tyr Ser Pro Phe Ala Gly Arg Leu Arg Asn Lys Glu
2204 65          70          75          80
2206 Asn Gly Asp Leu Glu Val Glu Cys Ser Gly Glu Gly Ala Val Phe Val
2207          85          90          95
2209 Glu Ala Met Ala Asp Asn Glu Leu Ser Val Leu Gln Asp Leu Asp Glu
2210          100          105          110
2212 Tyr Cys Thr Ser Leu Lys Gln Leu Ile Phe Thr Val Pro Met Asp Thr
2213          115          120          125
2215 Lys Ile Glu Asp Leu His Leu Leu Ser Val Gln Val Thr Ser Phe Thr
2216          130          135          140
2218 Cys Gly Gly Phe Val Val Gly Ile Ser Phe Tyr His Thr Ile Cys Asp
2219 145          150          155          160
2221 Gly Lys Gly Leu Gly Gln Phe Leu Gln Gly Met Ser Glu Ile Ser Lys
2222          165          170          175
2224 Gly Ala Phe Lys Pro Ser Leu Glu Pro Val Trp Asn Arg Glu Met Val
2225          180          185          190
2227 Lys Pro Glu His Leu Met Phe Leu Gln Phe Asn Asn Phe Glu Phe Val
2228          195          200          205
2230 Pro His Pro Leu Lys Phe Lys Lys Ile Val Lys Ala Ser Ile Glu Ile
2231          210          215          220
2233 Asn Phe Glu Thr Ile Asn Cys Phe Lys Gln Cys Met Met Glu Glu Cys
2234 225          230          235          240
2236 Lys Glu Asn Phe Ser Thr Phe Glu Ile Val Ala Ala Leu Ile Trp Leu
2237          245          250          255
2239 Ala Lys Thr Lys Ser Phe Gln Ile Pro Asp Ser Glu Asn Val Lys Leu
2240          260          265          270
2242 Met Phe Ala Val Asp Met Arg Thr Ser Phe Asp Pro Pro Leu Pro Lys

```



## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,570

DATE: 06/15/2001

TIME: 16:35:41

Input Set : A:\53679.APP.txt

Output Set: N:\CRF3\06152001\I866570.raw

```

2243          275          280          285
2245 Gly Tyr Tyr Gly Asn Val Ile Gly Ile Ala Gly Ala Ile Asp Asn Val
2246          290          295          300
2248 Lys Glu Leu Leu Ser Gly Ser Ile Leu Arg Ala Leu Ile Ile Ile Gln
2249 305          310          315          320
2251 Lys Thr Ile Phe Ser Leu Lys Asp Asn Phe Ile Ser Arg Arg Leu Met
2252          325          330          335
2254 Lys Pro Ser Thr Leu Asp Val Asn Met Lys His Glu Asn Val Val Leu
2255          340          345          350
2257 Leu Gly Asp Trp Arg Asn Leu Gly Tyr Tyr Glu Ala Asp Cys Gly Cys
2258          355          360          365
2260 Gly Asn Leu Ser Asn Val Ile Pro Met Asp Gln Gln Ile Glu His Glu
2261          370          375          380
2263 Ser Pro Val Gln Ser Arg Phe Met Leu Leu Arg Ser Ser Lys Asn Met
2264 385          390          395          400
2266 Gln Asn Gly Ile Lys Ile Leu Met Ser Met Pro Glu Ser Met Ala Lys
2267          405          410          415
2269 Pro Phe Lys Ser Glu Met Lys Phe Thr Ile Lys Lys Tyr Val Thr Gly
2270          420          425          430
2272 Ala Cys Phe Ser Glu Leu
2273          435
E--> 2280 (2)

```

*delete*

*FSI*

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

09/866,570 4

<210> 28

<211> 439

<212> PRT

<213> Artificial Sequence

see item 11 on Ena Summary Sheet

<400> 28

Met Glu Lys Thr Asp Leu His Val Asn Leu Ile Glu Lys Val Met Val

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/866,570

DATE: 06/15/2001

TIME: 16:35:42

Input Set : A:\53679.APP.txt

Output Set: N:\CRF3\06152001\I866570.raw

L:9 M:270 C: Current Application Number differs, Replaced Application Number  
 L:10 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
 L:787 M:340 E: (46) "n" or "Xaa" used: Feature required, for SEQ ID#:18  
 L:1233 M:258 W: Mandatory Feature missing, <220> FEATURE:  
 L:1233 M:258 W: Mandatory Feature missing, <223> OTHER INFORMATION:  
 L:1398 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:34  
 L:1398 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:34  
 L:1398 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:34  
 L:1409 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:35  
 L:1409 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:35  
 L:1409 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:35  
 L:1420 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:36  
 L:1420 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:36  
 L:1420 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:36  
 L:1431 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:37  
 L:1431 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:37  
 L:1431 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:37  
 L:1442 M:258 W: Mandatory Feature missing, <221> not found for SEQ ID#:38  
 L:1442 M:258 W: Mandatory Feature missing, <222> not found for SEQ ID#:38  
 L:1442 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:38  
 L:1665 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:48  
 L:2280 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:58